

# UCLA

## UCLA Previously Published Works

### Title

Multi-ancestry phenome-wide association of complement component 4 variation with psychiatric and brain phenotypes in youth

### Permalink

<https://escholarship.org/uc/item/6qx3b949>

### Journal

Genome Biology, 24(1)

### ISSN

1474-760X

### Authors

Hernandez, Leanna M

Kim, Minsoo

Zhang, Pan

et al.

### Publication Date

2023

### DOI

10.1186/s13059-023-02878-0

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>


Peer reviewed

RESEARCH

Open Access



# Multi-ancestry phenome-wide association of complement component 4 variation with psychiatric and brain phenotypes in youth

Leanna M. Hernandez<sup>1\*</sup>, Minsoo Kim<sup>1,2,3</sup>, Pan Zhang<sup>1</sup>, Richard A. I. Bethlehem<sup>4</sup>, Gil Hoftman<sup>1</sup>, Robert Loughnan<sup>5</sup>, Diana Smith<sup>5</sup>, Susan Y. Bookheimer<sup>1</sup>, Chun Chieh Fan<sup>5</sup>, Carrie E. Bearden<sup>1</sup>, Wesley K. Thompson<sup>5</sup> and Michael J. Gandal<sup>1,2,6,7,8\*</sup> 

\*Correspondence:  
leannahernandez@ucla.edu;  
Michael.gandal@penncmedicine.  
UPenn.edu

<sup>1</sup> Department of Psychiatry,  
Semel Institute, David Geffen  
School of Medicine, University  
of California, Los Angeles, Los  
Angeles, CA 90095, USA

<sup>7</sup> Lifespan Brain Institute at Penn  
Med and the Children's Hospital  
of Philadelphia, Philadelphia,  
PA, USA

Full list of author information is  
available at the end of the article

## Abstract

**Background:** Increased expression of the complement component 4A (*C4A*) gene is associated with a greater lifetime risk of schizophrenia. In the brain, *C4A* is involved in synaptic pruning; yet, it remains unclear the extent to which upregulation of *C4A* alters brain development or is associated with the risk for psychotic symptoms in childhood. Here, we perform a multi-ancestry phenome-wide association study in 7789 children aged 9–12 years to examine the relationship between genetically regulated expression (GREx) of *C4A*, childhood brain structure, cognition, and psychiatric symptoms.

**Results:** While *C4A* GREx is not related to childhood psychotic experiences, cognition, or global measures of brain structure, it is associated with a localized reduction in regional surface area (SA) of the entorhinal cortex. Furthermore, we show that reduced entorhinal cortex SA at 9–10 years predicts a greater number and severity of psychosis-like events at 1-year and 2-year follow-up time points. We also demonstrate that the effects of *C4A* on the entorhinal cortex are independent of genome-wide polygenic risk for schizophrenia.

**Conclusions:** Our results suggest neurodevelopmental effects of *C4A* on childhood medial temporal lobe structure, which may serve as a biomarker for schizophrenia risk prior to symptom onset.

**Keywords:** Schizophrenia, Psychosis, Neuroimaging, Genetics, Gene expression, Complement, Brain

## Background

Schizophrenia is a highly heritable neurodevelopmental disorder characterized by the presence of hallucinations, delusions, disorganized behavior, and an array of negative symptoms (e.g., flattened affect, reduced motivation) and cognitive deficits. While clinical diagnosis typically occurs in late adolescence or early adulthood, frank psychosis is



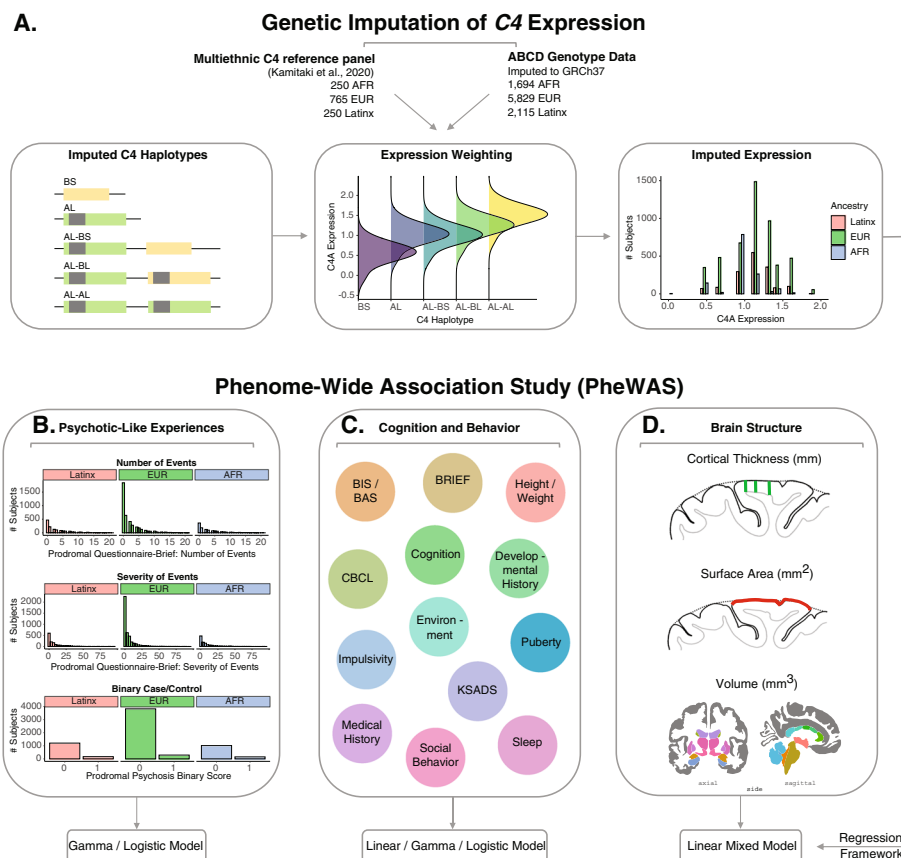
© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

often preceded by a prodromal period of weeks to years during which gradual changes in cognition, perception, and motivation occur [1]. An estimated 10% of youth between the ages of 9 and 18 experience subclinical psychosis-like experiences (PLEs), including atypical sensory experiences and delusional beliefs [2], which are associated with significantly greater odds of developing schizophrenia in adulthood [2, 3]. One proposed mechanism through which normative developmental processes may contribute to schizophrenia pathophysiology is through atypical brain development and refinement of synaptic connections across childhood and adolescence [4, 5]. A major goal of modern psychiatric research is to identify the genetic and brain-based antecedents of prodromal psychosis, with the ultimate goal of facilitating early intervention, prevention, and better treatment.

The most statistically significant genetic association with schizophrenia is located within the major histocompatibility complex (MHC) on chromosome 6—a region comprised of genes critical to innate and adaptive immunity [6–10]. Subsequent fine mapping of this locus revealed that a significant proportion of this signal can be attributed to multiallelic copy number variation of the complement component 4A (*C4A*) gene [11]. The complement system plays a crucial role in the peripheral immune response [12], and in the brain, it also plays an active role in pruning and facilitating microglial-mediated engulfment of synapses during development [13–16]. These findings have led to the hypothesis that the overactivation of complement genes in the brain may contribute to altered neural development and synaptic pruning, ultimately conferring risk for the phenotypic expression of schizophrenia symptoms.

The *C4* locus encodes for *C4A* and *C4B* genes, which show common but complex structural variation. Each human haplotype carries a range of *C4A* and/or *C4B* copies, present in either long or short forms (i.e., *C4AL*, *C4AS*, *C4BL*, *C4BS*) depending on the presence of a human endogenous retroviral insertion (HERV) in intron 9 [17]. A higher *C4A* copy number is associated with a greater risk for schizophrenia, proportional to its effects on *C4A* expression [11, 18, 19]. *C4A* copy number also shows sex-dependent effects on schizophrenia risk, with stronger effects in males relative to females [18]. Importantly, while recent neuroimaging work suggests that *C4A* also affects cognition and brain structure in the general adult population [20], no study to date has examined the impact of *C4A* on childhood psychosis-related symptoms or brain development.

Here, we performed a comprehensive, phenome-wide association study between genetically regulated expression (GREx) at the *C4* locus and childhood psychiatric, behavioral, and brain structural phenotypes to elucidate the role of *C4* genetic variation on human development (Fig. 1). *C4A* GREx was imputed into a large, ancestrally diverse sample of 9–10-year-old youth who participated in the Adolescent Brain Cognitive Development<sup>SM</sup> (ABCD) Study [21]. We then assessed whether *C4A* GREx was associated with psychosis-like experiences as well as longitudinal psychiatric and cognitive phenotypes. We find that *C4A* GREx is not broadly related to childhood psychiatric symptoms or cognition in 9–10-year-old youth but is associated with a reduction in entorhinal cortex surface area. Furthermore, we find that entorhinal cortex surface area at 9–10 years of age is predictive of the number and severity of psychotic-like events at 1- and 2-year follow-up visits. Together, this phenome-wide characterization of the



**Fig. 1** Study design. **A** Individual-level genotype data from ABCD subjects was used to impute *C4* structural alleles using a multi-ancestry reference panel (Kamitaki et al., [18]). *C4A* and *C4B* brain GREx was calculated using previously described weights (Sekar et al., [11]). **B** Generalized linear models were used to test for the associations between *C4A* GREx with psychotic-like experiences (PLE) in ABCD youth. Associations were examined in the whole multi-ancestry cohort, as well as within ancestry- and sex-specific subgroups. **C** We performed a phenome-wide association study (PheWAS) to assess the associations between *C4A* GREx and a host of developmental, cognitive, and behavioral phenotypes (the “Methods” section). **D** Linear mixed-effects models tested the association between predicted *C4A* GREx and neuroimaging measures of cortical thickness, surface area, and volume

developmental effects of *C4A* provides evidence for regional effects on temporal lobe structure, which may be an early biomarker for schizophrenia risk.

## Results

### Genetically regulated expression of *C4A* in the multi-ancestry ABCD youth cohort

Genetic and behavioral data were obtained from a community sample of sociodemographically and ethnically diverse youth who participated in the ABCD Study, an ongoing longitudinal investigation of child development in the USA [22]. Following quality control, including the removal of subjects with low *C4* imputation quality, the sample consisted of 9638 ancestrally diverse youth (1694 African ancestry [AFR]; 5829 European ancestry [EUR]; 2115 Latinx ancestry; Fig. 1A). Imputation accuracy was confirmed by comparing the proportion of youth assigned to each common *C4* structural allele to rates previously reported in a large sample of 765 EUR and 250 AFR ancestry individuals, from which the imputation panel was derived [18] (Additional file 1:

Table S1). Subsequent analyses were restricted to 7789 youth with the 5 most common *C4* structural haplotypes (1327 AFR; 4919 EUR; 1543 Latinx; 47.7% female; Additional file 1: Table S2), although in sensitivity analyses, results were concordant across the full spectrum of haplotypes. *C4* structural haplotypes were then used to calculate GREx for *C4A* and *C4B* using previously described weights [11]. We validated the accuracy of *C4* imputation weights for predicting gene expression in ancestrally diverse cohorts using data from 340 AFR, 863 EUR, and 50 Latinx donors from the PsychENCODE Consortium (Additional file 2: Fig. S1). We further validated that these weights were accurate in the subset of 5–15-year-old samples from PsychENCODE, which match the adolescent period investigated in this study. As expected, imputed expression values for *C4A* and *C4B* were negatively correlated ( $R = -0.25$ ,  $P < 2.2e-16$ ; Additional file 2: Fig. S2). To control for this, downstream statistical analyses included GREx for both *C4A* and *C4B*.

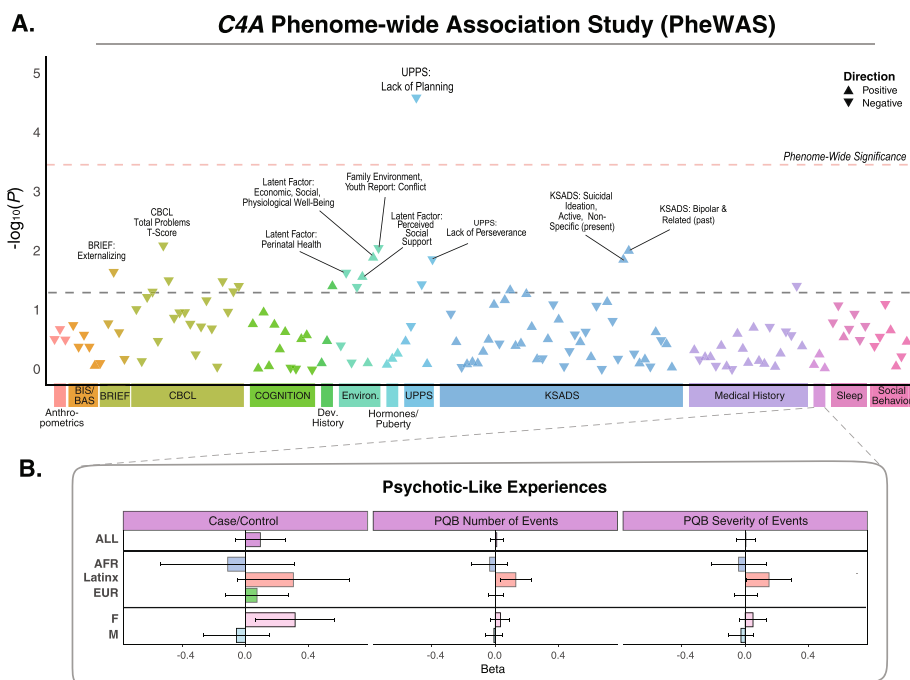
#### Adequacy of statistical power to detect GREx-phenotype associations

Power analyses were conducted in G\*Power [23] to determine the range of detectable effects with the current sample size. With an estimated  $N$  of 6500, alpha of 0.05, 1 tested predictor, and 11 covariates, we have 95% power to detect effects (Cohen's  $f^2$ ) as small as 0.002. As previous research in adults suggests that we can expect to see a significant association between *C4A* GREx and cognitive measures at Cohen's  $f^2 = 0.056$  [20], power calculations indicate that our study is sufficiently powered to detect small but statistically significant effects, equivalent to a Cohen's  $d$  of 0.004.

#### No widespread *C4A* associations with psychotic-like experiences or childhood cognition

Models testing associations between *C4A* GREx with rates and severity of PLEs were restricted to unrelated individuals who had complete data for the Prodromal Questionnaire Brief Version [24] (PQ-B) and the Kiddie Schedule for Affective Disorders and Schizophrenia [25] (KSADS-5) ( $N = 6655$ ; 1168 AFR; 4113 EUR; 1374 Latinx). As previous research in clinical high-risk samples indicates that suspiciousness and unusual thought content best predict conversion to psychosis over a 2-year period [26], a binary prodromal psychosis score ( $PP_{bin}$ ) was computed by categorizing youth who endorsed experiencing at least 3 PLEs reflecting suspicious or unusual thought content and significant distress on the PQ-B and whose parents indicated the presence of hallucinations or delusions on the KSADS-5 ( $N = 611$  cases,  $N = 6044$  controls). We did not observe any significant associations between *C4A* GREx and psychosis symptoms (i.e.,  $PQ-B_{sym}$ ,  $PQ-B_{sev}$ ,  $PP_{bin}$ ; Fig. 2, Additional file 1: Table S3, Additional file 2: Figs. S3-S5). As previous work has identified sexually dimorphic effects of *C4A* on the brain and behavior [18, 19], we further examined a potential interaction between sex and *C4A* GREx—which was not significant for any of the psychosis variables tested (Additional file 1: Table S3, Additional file 2: Figs. S6-S8).

In adults and schizophrenia patients, genetic variation within the *C4* locus, as well as in genes that regulate the complement system, has been associated with poorer cognition and memory [20, 28–30]. To assess whether similar effects are present in childhood, we performed a phenome-wide association study (PheWAS) using data from 145 cognitive and psychiatric phenotypes available from the ABCD dataset. In contrast to the findings in adults, *C4A* GREx was not robustly associated with childhood cognition or



**Fig. 2** Phenome-wide association between *C4A* GREx and cognitive, behavioral phenotypes in the ABCD cohort. **A** The relationship between *C4A* GREx and 145 phenotypes was tested in the multi-ancestry cohort (*N*'s per phenotype ~ 6000). Increased genetically regulated expression of *C4A* was associated with lower scores on the UPPS-P Impulsive Behavior Scale (Lynam, [27]) lack of premeditation/planning subscale (FDR-corrected  $P = 0.004$ ). Phenotypes are grouped into broad categories by color. Negative or positive associations with *C4A* GREx are indicated by the direction of arrows. The FDR-based threshold for phenome-wide significance is indicated by the red dashed line. **B** Association (absolute  $\beta \pm$  standard error) between *C4A* GREx and each of the three psychosis variables investigated (binary case/control, PQ-B number PQ-B severity) in the multi-ancestry cohort (i.e., "ALL" youth [as in top panel]), as well as within each ancestry and sex separately. See Additional file 1: Table S3 for the regression model summary statistics. The results by ancestry and sex are shown in Additional file 2: Figs. S3-S8

psychiatric phenotypes (Fig. 2, Additional file 1: Table S3, Additional file 2: Figs. S3-S8). The only association that surpassed correction for multiple testing in the multi-ancestry cohort was between *C4A* GREx and the UPPS-P Impulsive Behavior Scale [27] lack of premeditation/planning subscale ( $\beta = -0.05$ , 95% CI  $[-0.07, -0.04]$ , FDR-corrected  $P < 0.004$ ), suggesting that youth with higher *C4A* GREx show less impulsivity and more careful thinking before taking action. This effect was partially mediated by entorhinal cortex size (indirect effect  $-0.014$ ,  $P = 0.006$ ).

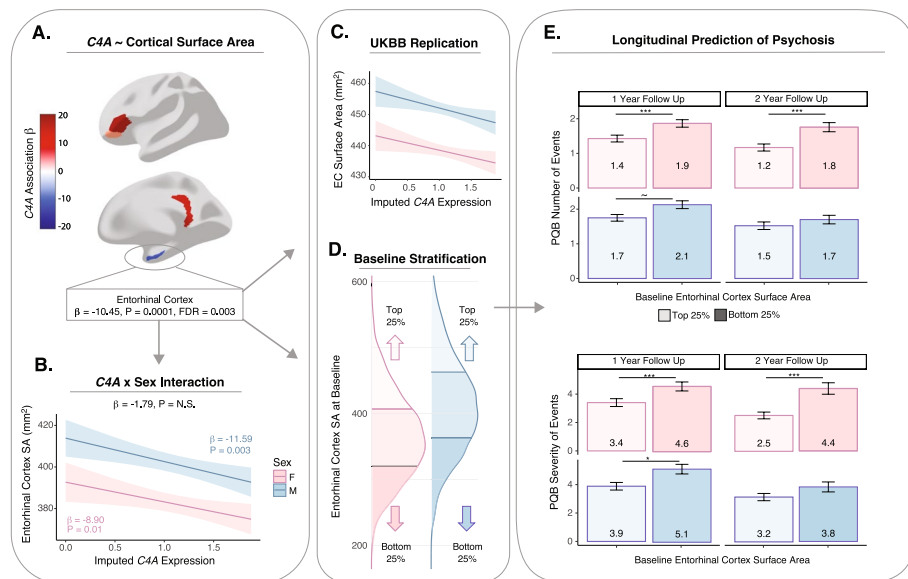
Additional PheWAS were performed using available data from follow-up time points 1 and 2 years later. There were no FDR-significant associations between *C4A* GREx and PheWAS variables at the 1- or 2-year follow-up (Additional file 1: Tables S4-S5, Additional file 2: Figs. S9-S14). Finally, to assess the convergence between *C4* transcriptomic and genetic associations, regression analyses were performed for the joint model assessing the impact of 14 unique *C4* allelic combinations on childhood PLEs. The *C4* AL:AL haplotype showed nominally significant associations with greater PLEs; however, after Bonferroni correction, the results were no longer significant (Additional file 2: Fig. S15).

### Replicable effects of the *C4* locus on the entorhinal cortex structure

Our PheWAS results above indicate limited detectable effects of *C4* variation on cognitive or psychiatric outcomes in youth. However, in schizophrenia, subtle brain structural changes are known to precede clinical and behavior manifestations [31–33]. To determine whether *C4A* affects childhood brain development, we next tested the association between GREx and measures of cortical thickness (CT), surface area (SA), and volume (VOL) across the whole brain as well as within regional parcellations using the Desikan-Kiliany atlas [34].

In the ancestrally diverse ABCD cohort, *C4A* GREx was associated with reduced regional SA of the entorhinal cortex ( $\beta = -10.45$ , 95% CI  $[-13.13, -7.76]$ , FDR-corrected  $P = 0.003$ ; Fig. 3A, Additional file 1: Table S6). Furthermore, we found that the relationship between *C4A* GREx and entorhinal cortex was significant for both female and male youths (female:  $\beta = -8.90$ , 95% CI  $[-12.46, -5.34]$ ,  $P = 0.01$ ; male:  $\beta = -11.6$ , 95% CI  $[-15.56, -7.63]$ ,  $P = 0.003$ ; Fig. 3B). No other brain structure associations were observed after correction for multiple testing (Additional file 1: Tables S7–S9).

To assess the robustness and reproducibility of the impact of *C4A* GREx on entorhinal cortex brain structure, we performed analogous analyses in 16,147 EUR ancestry



**Fig. 3** *C4A* GREx is associated with entorhinal cortex surface area and predicts longitudinal psychosis symptoms. **A** Linear regression analyses were performed in the multi-ancestry cohort to test the association between *C4A* GREx and regional brain structure ( $N$ 's with brain imaging data passing quality control  $\sim 6500$ ). Brain regions demonstrating a nominally significant association between *C4A* GREx and surface area ( $\text{mm}^2$ ) are shown in color ( $P < 0.05$ ); only the association with the entorhinal cortex survived FDR correction. The results for all brain regions are provided in Additional File 1: Tables S6–S9. **B** Female and male youths showed a significant effect of *C4A* GREx on reduced entorhinal cortex surface area. **C** Replication of the relationship between *C4A* GREx and entorhinal cortex (EC) surface area in 8357 adult females and 7790 adult males of European ancestry from the UK Biobank. **D** Youths falling in the top and bottom quartiles in terms of baseline entorhinal surface area were identified. **E** Smaller entorhinal cortex surface area (bottom quartile) at baseline predicted greater number and severity of psychosis-like experiences at the 1- and 2-year follow-up time points. The mean number and severity of PLEs are shown inside each bar plot.  $\sim P = 0.05$ ,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$

subjects from the UK Biobank [35]. In the UK Biobank, *C4A* GREx was also significantly associated with entorhinal cortex SA, with the same direction of effect (full sample:  $\beta = -5.01$ , 95% CI  $[-7.95, -2.06]$ ,  $P = 0.00087$ ; female  $\beta = -5.47$ , 95% CI  $[-8.29, -0.85]$ ,  $P = 0.02$ ; male  $\beta = -4.57$ , 95% CI  $[-10.10, -0.84]$ ,  $P = 0.02$ ; Fig. 3C).

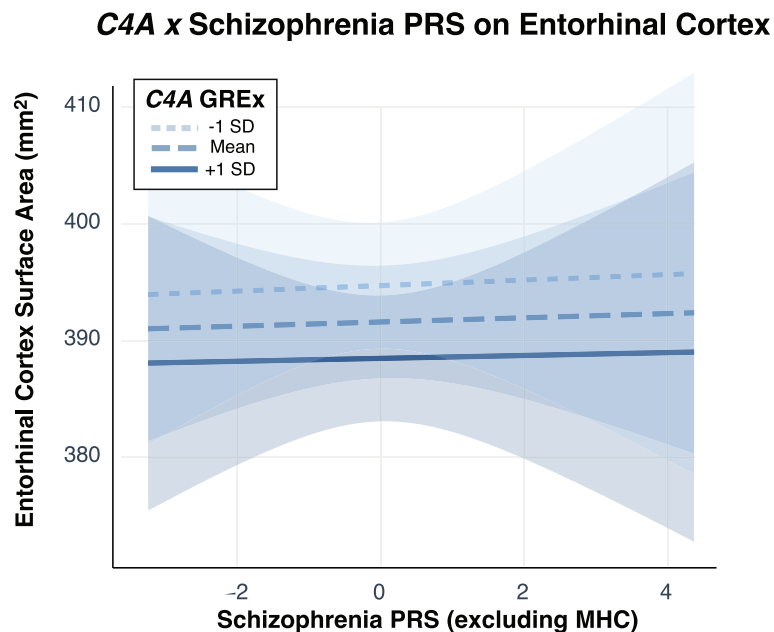
The previous analyses show that *C4A* GREx is associated with a localized reduction in entorhinal cortex SA. In psychosis spectrum patients, the size of the entorhinal cortex is a predictor of the severity of delusions and negative symptoms [36, 37]. To assess this, we leveraged longitudinal psychosis data available from the ABCD 1- and 2-year follow-up visits. Due to the differences in overall brain size between male and female youths, analyses were performed in each sex separately and focused on comparing youth in the top and bottom quartile for baseline entorhinal cortex SA (Fig. 3D). We observed a significant difference in both the number and severity of reported PLEs at the 1-year follow-up when comparing between youth in the top and bottom quartiles for entorhinal cortex SA; more specifically, female and male youths with smaller entorhinal cortex SA at baseline reported greater frequency of PLEs and more distressing PLEs 1-year later (Wilcoxon Rank Sum Tests  $r = 0.05$ – $0.09$ ,  $P = 0.0002$ – $0.04$ ; Fig. 3E, Additional file 1: Table S10). At the 2-year follow-up, female youth showed significant differences in PLEs as a function of baseline entorhinal cortex SA; here again, smaller entorhinal cortex SA at baseline was associated with greater frequency and severity of PLEs ( $r = 0.11$ – $0.12$ ,  $P = 0.0002$ ; Fig. 3E, Additional file 1: Table S10). There were no significant associations between entorhinal cortex SA and PLEs in male youth at the 2-year follow-up. Parallel regression analyses modeling entorhinal cortex SA as a continuous predictor were significant in females only, such that smaller entorhinal cortex at baseline predicted greater frequency and severity of PLEs at the 2-year follow-up (PQ- $B_{\text{sym}}$   $\beta = -0.001$ , 95% CI  $[-0.002, -0.0001]$ ,  $P = 0.03$ ; PQ- $B_{\text{sev}}$   $\beta = -0.002$ , 95% CI  $[-0.003, -0.0005]$ ,  $P = 0.004$ ).

#### Effects of *C4A* on the entorhinal cortex are independent of polygenic risk for schizophrenia

Schizophrenia is highly polygenic with thousands of genetic variants of small effect size acting additively to confer risk. To better understand the relationship between genetic risk for schizophrenia at the *C4* locus and additive polygenic risk across the genome (outside the *C4* locus), we computed polygenic risk scores (PRS) in ABCD youth of EUR ancestry ( $N = 3730$  unrelated subjects) using the genome-wide association summary statistics of schizophrenia (40,675 cases, 64,643 controls [10]), excluding variants in the MHC region. Similar to previous findings [38], we did not observe a significant association between schizophrenia PRS and psychotic symptoms in youth—nor did we observe an interaction between *C4A* GREx and schizophrenia PRS in predicting PQ- $B_{\text{sym}}$ , PQ- $B_{\text{sev}}$ , or PP $_{\text{bin}}$  in ABCD youth of EUR ancestry or when examining EUR females and males separately (Additional file 1: Table S11).

Finally, we tested whether schizophrenia PRS—or its interaction with *C4A* GREx—was a significant predictor of brain structure. Surprisingly, schizophrenia PRS showed no association with entorhinal cortex structure, nor was there an interaction with *C4A* GREx, indicating a dissociable effect of complement activation and schizophrenia polygenic burden on this brain structure (Fig. 4, Additional file 1: Tables S11–S14). To





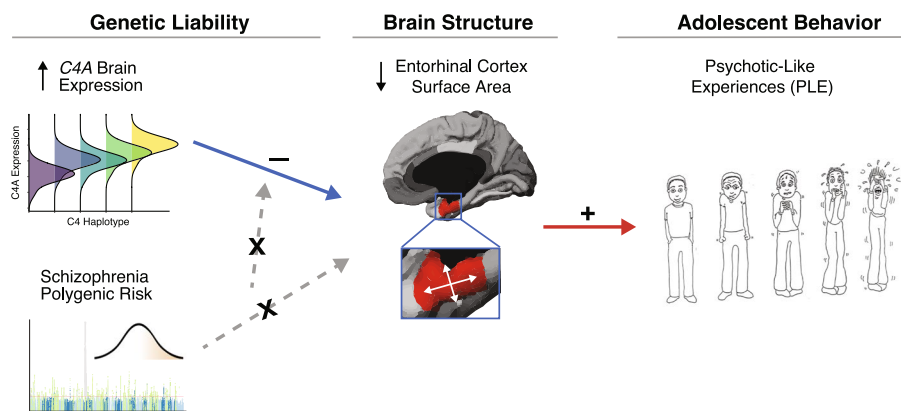
**Fig. 4** Interactions between *C4A* GREx and polygenic risk for schizophrenia. The interaction between *C4A* GREx and schizophrenia PRS was not a significant predictor of entorhinal cortex surface area in 9–10-year-old youth of European ancestry ( $N = 3349$ ); higher *C4A* GREx was associated with smaller entorhinal cortex surface area across all levels of PRS

confirm this, the association between *C4A* GREx and entorhinal cortex SA remained significant after controlling for schizophrenia PRS ( $\beta_{C4A} = -8.88$ , 95% CI  $[-12.46, -.29]$ ,  $P = 0.01$ ). We also tested whether the interaction between brain surface area polygenic scores and *C4A* GREx predicted entorhinal cortex SA; the interaction was not significant ( $P > 0.05$ ). Together, these findings suggest that the effects of *C4A* GREx on entorhinal cortex SA are specific to the *C4* locus and are distinct from the effects of genetic liability for schizophrenia and adult brain size.

## Discussion

Here, we conduct a comprehensive multi-ancestry, phenome-wide association study of complement component 4 (*C4*) variation with childhood cognitive, psychiatric, and neuroimaging phenotypes. We find that genetically regulated expression of *C4A* is not associated with broad psychiatric symptoms or brain structure in 9–10-year-old children, but rather shows localized and replicable effects on the reduced regional surface area of the entorhinal cortex. Importantly, this *C4A*-mediated reduction in the entorhinal cortex surface area at 9–10 years of age predicted higher rates and greater severity of psychosis-like experiences at follow-up time points 1 and 2 years later, indicating that the effects of *C4A* on temporal lobe structure may be a biomarker for schizophrenia risk prior to symptom onset (Fig. 5).

Despite having sufficient power to detect associations with small effect sizes, we find that predicted *C4A* expression is not robustly associated with psychiatric symptoms or cognition in 9–10-year-old youth. These null results are surprising given the strength of the association between *C4A* and schizophrenia risk [11]. However, these negative findings comport with recent studies assessing the extent to which polygenic risk for



**Fig. 5** Model for the effects of *C4A* brain expression on childhood brain structure. The effects of increased *C4A* GREx on reduced entorhinal cortex volume in childhood are dissociated from the effects of genome-wide polygenic risk for schizophrenia. Reduced entorhinal cortex surface area in childhood is associated with higher rates of youth-reported psychotic-like experiences (PLEs) in early adolescence, as well as greater severity of events. *Note:* Image under “psychotic-like experiences” taken from the Prodromal Questionnaire Brief Version (PQ-B) to assess the severity of distress associated with psychotic-like events

schizophrenia is associated with broad psychiatric phenotypes in the ABCD cohort, which have found limited associations with fluid cognition and an index of binary psychosis severity in ABCD youth of European ancestry [39, 40]. Nevertheless, we identified a significant inverse association between predicted *C4A* expression and lack of planning on the UPPS-P Impulsive Behavior Scale [27]. Higher scores on this subscale are indicative of a tendency not to plan ahead; thus, we find that *C4A* brain expression is related to less severe symptomatology. The effects of schizophrenia genetic risk at the *C4* locus and across the genome may exert more extensive effects on cognition and behavior during later developmental periods (e.g., adolescence and early adulthood), when clinically significant features of psychosis and schizophrenia are more likely to occur. We speculate that the developmental timing and tempo of synaptic pruning may be an important contributor to the emergence of psychiatric symptoms, such that increased pruning during early childhood may be developmentally advantageous until reaching a threshold of “over” pruning later in life.

In addition to the robust genetic association signal observed between schizophrenia and the major histocompatibility complex (MHC) locus, this region also shows associations with depression [41, 42] and to a weaker extent bipolar disorder [7, 43]. However, in both depression and bipolar disorder, dissecting the MHC signal using fine mapping techniques suggests that the observed associations are not driven by genetic variation in *C4* haplotypes and arise from other variants in the region [44, 45]. Our null phenome-wide association study (PheWAS) testing the effects of the *C4* locus on childhood psychiatric phenotypes agree these findings, suggesting that other sources of genetic variation apart from the *C4* locus are likely influencing childhood cognition and behavior. Further large-scale studies will be needed to disentangle the genetic signal between MHC locus and diverse psychiatric disorders, and to examine the effects of other MHC-associated variants on neurodevelopmental trajectories.

In this demographically diverse multi-ancestry cohort, we observed a regionally localized association between *C4A* GREx and surface area of the entorhinal cortex. Our

entorhinal cortex findings are consistent with the extant literature implicating medial temporal lobe structures in the pathophysiology of schizophrenia [46–48]. Indeed, in vivo neuroimaging has consistently found atypical volume of medial temporal lobe structures in patients with psychotic disorders relative to controls [49–51], and several postmortem studies have found aberrant entorhinal cortex cytoarchitecture in schizophrenia patients, suggestive of either atypical neuronal migration or altered pruning during neurodevelopment [52–56].

The entorhinal cortex serves as a relay for corticohippocampal interactions, serving as the major input and output structure of the hippocampus [57]. Recent high-field magnetic resonance imaging (MRI) research in humans suggests an anteroposterior division of entorhinal cortex functional connectivity, with anterior regions connected to anterior temporal brain areas critical for recognition memory, emotional processing, and social cognition and posterior regions connected to posterior medial brain areas that support episodic memory and spatial navigation [58, 59]. Because the entorhinal cortex receives afferent projections from multiple cortical and subcortical brain regions and is the primary source of information to the hippocampus [60], disruptions to this region during development have the potential to affect both the integration of sensory information and memory [58, 61, 62]. In line with this hypothesis, previous work has shown that smaller entorhinal cortex volume is a predictor of cognition and negative/disorganized symptoms in psychosis spectrum youth [37]. Similarly, we find that reduced entorhinal cortex surface area at 9–10 years of age predicts greater number and severity of psychosis-like events 1 and 2 years later. As brain volume is a function of surface area and cortical thickness, our findings of smaller surface area in youth with high genetic risk for schizophrenia at the *C4* locus help to more precisely identify potential morphometric biomarkers predictive of psychosis risk.

Our *surface area*-specific findings in preadolescent youth contrast with adult neuroimaging studies, which have typically found more widespread reductions in *cortical thickness* in schizophrenia patients [63] and in clinically high-risk individuals who transition to psychosis [31]—although we did replicate the effects of *C4A* GREx on entorhinal cortex surface area in adults subjects from the UKBiobank. These contradictory findings may be explained by developmental differences in structural neuroimaging parameters. Recent neuroimaging studies have shown that cortical surface area continues to expand throughout the postnatal period and early childhood [64], while adolescence and early adulthood are characterized by nonlinear decreases in cortical thickness [65], likely driven by developmentally regulated alterations in cortical myelination and pruning. Thus, genetic risk for schizophrenia may have distinct effects on diverse aspects of brain structure at different developmental periods. Importantly, *C4A* cis-eQTLs (i.e., genomic loci that explain variation in gene expression) have been identified in the fetal brain [66], suggesting that the effects of *C4A* GREx on brain morphology may begin early in development. We therefore hypothesize that the observed effects of *C4A* GREx on the cross-sectional entorhinal cortex surface area in children stem from early alterations in brain development and that future studies of the effects of *C4A* on brain development are likely to observe associations with morphological metrics in flux within a particular developmental window (e.g., adolescent cortical thickness).

Notably, the effects of predicted *C4A* GREx on entorhinal cortex surface area are independent and dissociable from genome-wide polygenic risk for schizophrenia. We provide several lines of evidence to support this conclusion. First, polygenic risk scores (PRS) trained using weights from the largest available GWAS of schizophrenia [10] failed to predict entorhinal cortex surface area in youth of European ancestry. Second, when examining the interaction between *C4A* GREx and PRS, we find no evidence that schizophrenia PRS moderates the relationship between *C4A* GREx and entorhinal cortex surface area, and finally, the observed relationship between *C4A* GREx and entorhinal cortex surface area remains significant even when controlling for the effects of PRS. These findings indicate that the observed effects on the entorhinal cortex surface area are exclusive to genetic risk for schizophrenia at the *C4* locus and are distinct from the effects of schizophrenia polygenic risk.

This study has several important limitations. First, our analyses were restricted to youth within the early to mid-adolescent age range. It remains unknown the extent to which *C4A* GREx may affect neurodevelopment during infancy and very early childhood, as well as during adolescence, when many disorders of mental health—including schizophrenia—begin to emerge. Second, our PheWAS analyses were limited to a single neuroimaging modality, and it is unclear whether genetic risk for schizophrenia on the *C4* locus may impact measures of brain functional and structural connectivity, as well as the developmental timing with which these associations may occur. Third, while there are many environmental variables that are likely to influence brain and behavioral development, the current study did not examine the relationship between *C4A* GREx and environmental factors known to increase the risk for schizophrenia. It will be crucial for future studies to examine the impact of *C4A* expression on trajectories of brain development using multimodal neuroimaging techniques and incorporating potential interactions with environmental moderators.

## Conclusions

In sum, the current study reports the first multi-ancestry phenome-wide association of genetic variation with the *C4* locus in children. Our findings indicate that increased *C4A* brain expression in youth is associated with a localized reduction in the size of the entorhinal cortex that is independent of polygenic risk for schizophrenia. Overall, these data help to identify genetically driven brain-based biomarkers predictive of psychosis risk. Ultimately, a comprehensive understanding of the genetic and environmental precursors to psychosis will require longitudinal indices of brain development across the transition to adolescence and young adulthood, when clinically significant behavioral manifestations of psychiatric disorders often begin to emerge.

## Methods

### Participants

Data were obtained from the Adolescent Brain and Cognitive Development (ABCD) Study, an ongoing longitudinal investigation of brain and behavioral development in 11,878 children in the United States of America (USA). Youth were recruited

primarily from schools within 21 catchment areas chosen to represent the ethnic and sociodemographic diversity of the larger US population [22]. Analyses were performed using de-identified phenotypic data from the baseline time point obtained from the ABCD 3.0 National Data Archive release (DOI 10.15154/1519007). Analyses were restricted to youth with available genetics data. See Additional file 1: Table S1 for subject characteristics.

### Genotyping and imputation

Genotyping for the ABCD dataset was performed at the Rutgers University Cell and DNA Repository using the Affymetrix NIDA Smokescreen Array (RUCDR). Quality control was performed by the RUCDR and the ABCD Data Analysis and Informatics Core (DAIC) and included filtering for call signals and rates, and processing through the Ricopili pipeline [67]. Genotyping data was merged with data from the 1000 Genomes Project, and ancestry was assigned using the  $k$ -nearest neighbors classification with the first four genetic principal components.

$C4$  structural alleles were imputed for the ABCD dataset using Beagle4.1 [68] with the latest multi-ancestry  $C4$  imputation reference panel composed of 1265 samples of diverse ancestry [18].  $C4A$  and  $C4B$  genetically regulated expression (GREx) was predicted using previously described expression weights [11].  $C4$  structural allele dosages were multiplied by the following coefficients to compute GREx:

$$C4A \text{ GREx} = (0.47 \times AL) + (0.47 \times AS) + (0.20 \times BL)$$

$$C4B \text{ GREx} = (1.03 \times BL) + (0.88 \times BS)$$

Downstream analyses were restricted to samples of African (AFR), European (EUR), and Latinx ancestries as the imputation panel was composed of those individuals, and  $C4$  imputation accuracy (based on probabilistic dosage) was lower for Southeast Asian and East Asian. Individuals with an average imputed probabilistic dosage  $< 0.7$  were excluded from subsequent analyses. To further ensure the use of high-quality imputed data, analyses were restricted to individuals with the five most common  $C4$  structural haplotypes (AL, AL-AL, AL-BL, AL-BS, BS) resulting in a final sample of 7789 youth (see Additional file 1: Table S2).

For the UK Biobank, chromosome 6 of the phased haplotype data (i.e., `ukb_hap_chr6_v2.bgen`) was used to impute  $C4$  structural alleles. Beagle4.1, with the multi-ancestry  $C4$  reference panel, was used for imputation, as was done for the ABCD dataset.

### Polygenic risk prediction

Polygenic risk scores (PRS) were calculated for EUR subjects. Genotype imputation was performed on the Michigan Imputation Server [69] using the TOPMed reference panel. Quality control included filtering for genotyping rate  $> 0.99$ , sample missingness  $< 0.01$ , Hardy-Weinberg Equilibrium  $P > 1 \times 10^{-6}$ , minor allele frequency  $> 1\%$ , and imputation score  $> 0.8$ . GCTA v1.93.31 [70] was used to compute the genetic relationship matrix; a relatedness cutoff of 0.05 was applied. Genetic variants in the

major histocompatibility complex (MHC) locus were excluded prior to PRS calculation. PRS were generated using summary statistics from the largest available genome-wide association study of schizophrenia [10] (40,675 cases, 64,643 controls) and adult brain surface area [71] (33,992 subjects) in PRS-CS [72], applying the LD reference panel from the 1000 Genomes Project phase 3 samples and the recommended global shrinkage parameter for highly polygenic traits ( $\phi = 1e-2$ ). Analyses assessing the interaction between *C4* expression and PRS were restricted to youth with common *C4* structural haplotypes (AL, AL-AL, AL-BL, AL-BS, BS), resulting in a final sample of 3730 EUR youth (female  $N = 1737$ ; male  $N = 1993$ ) and 7,715,663 genetic variants.

### Neuroimaging data acquisition and preprocessing

Details of the ABCD magnetic resonance imaging (MRI) protocol have been described previously [73]. Briefly, imaging data were collected on either a Siemens Prisma, Phillips, or GE 750 3T scanner using 32-channel head or 64-channel head/neck coils. T1-weighted scan parameters were as follows: Siemens - matrix size  $256 \times 256$ , 176 slices, FOV  $256 \times 256$ , resolution (mm)  $1 \times 1 \times 1$ , TR 2500 ms, TE 2.88 ms, flip angle  $8^\circ$ , total scan time 7:12; Phillips - matrix size  $256 \times 256$ , 225 slices, FOV  $256 \times 240$ , resolution (mm)  $1 \times 1 \times 1$ , TR 6.31 ms, TE 2.9 ms, flip angle  $8^\circ$ , total scan time 5:38; and GE - matrix size  $256 \times 256$ , 208 slices, FOV  $256 \times 256$ , resolution (mm)  $1 \times 1 \times 1$ , TR 2500 ms, TE 2 ms, flip angle  $8^\circ$ , total scan time 6:09.

Tabulated data representing T1-weighted measurements of subcortical volume (VOL), cortical thickness (CT), and surface area (SA) as derived from the Desikan parcellation atlas [34] in FreeSurfer v5.3 (<http://surfer.nmr.mgh.harvard.edu/>) were downloaded from the NDA. ABCD-recommended imaging inclusion criteria were followed to ensure the inclusion of high-quality MRI scans, and this excluded participants with serious MR findings and participants whose T1-weighted images or FreeSurfer parcellations failed to pass ABCD DAIC quality control. Regional MRI metrics were averaged across hemispheres.

### Phenotypic data

Psychotic experiences were assessed with the Prodromal Questionnaire Brief Version [24] (PQ-B) and the Kiddie Schedule for Affective Disorders and Schizophrenia [25] (KSADS-5). The PQ-B is a 21-item child-report questionnaire designed to measure the presence and severity of psychotic-like experiences (PLEs) in childhood. PQ-B symptom scores ( $PQ-B_{sym}$ ) were calculated as the total number of endorsed PLEs; severity score ( $PQ-B_{sev}$ ) was calculated as the child-reported level of distress (range 1–5) across all endorsed items. In line with previous research indicating that unusual thought content and suspiciousness are most predictive of psychosis risk [26], a binary score ( $PP_{bin}$ ) was developed by defining prodromal psychosis “cases” as youth who endorsed experiencing at least 3 PLEs reflecting suspicious or unusual thought content (i.e., PQ-B items 1, 4, 5, 7, 8, 11, 12, 13–18), significant distress related to these experiences (i.e., PQ-B distress scores  $> 6$  across items 1, 4, 5, 7, 8, 11, 12, 13–18), and whose parents indicated the child experienced hallucinations, delusions,

and associated psychotic symptoms on the KSADS-5 assessment; youth not meeting these criteria served as the “control” group. Cognitive and psychiatric variables from the reported *C4* phenome-wide association study (PheWAS) can be found in Additional file 1: Table S3.

Rates and severity of childhood PLEs were non-normally distributed, with the majority of youth reporting zero or few PLE symptoms ( $PQ-B_{sym}$ ;  $M = 2.55$ ,  $SD = 3.52$ ) and endorsing low levels of PLE severity ( $PQ-B_{sev}$ ;  $M = 6.09$ ,  $SD = 10.39$ ; Fig. 1B).

### Statistical analyses

Generalized linear models with gamma family and log link were run using the *glm* package in R version 3.6.3 to test the association between *C4* GREx and continuous measures of psychotic experiences (i.e.,  $PQ-B_{sym}$ ,  $PQ-B_{sev}$ ), as these variables demonstrated a right-skewed distribution. The association between *C4* GREx and  $PP_{bin}$  was tested via logistic regression with binomial family using the *glm* package in R. Fixed effects included predicted *C4A* and *C4B* GREx, age, socioeconomic status (SES; taken as the average of parent education and income at the baseline time point), four genetic principal components, sex, and site ID. Statistical models run in related individuals using family ID as a random effect failed to converge; thus, analyses were restricted to one randomly selected subject from each family.

Models testing multivariable phenome-wide associations were also restricted to one individual per family. For normally distributed variables, linear models were fitted, while for non-normally distributed variables, generalized linear models with gamma distribution and log link were fitted. *C4A* and *C4B* GREx, age, SES, four genetic principal components, sex, and site ID were specified as fixed effects. Benjamini-Hochberg false discovery rate (FDR) correction was used to account for the number of variables tested at each time point (i.e., baseline, 1-year follow-up, 2-year follow-up). The significance of covariates of interest was assessed using the likelihood ratio test.

To test the association between *C4* GREx and neuroimaging indices of brain volume, cortical thickness, and surface area, linear mixed models were run via the *lme4* package in R. *C4A* and *C4B* GREx, age, four genetic principal components, SES, and sex were entered as fixed effects; MRI device serial number and family ID were entered as random effects. To control for global brain effects, whole brain volume, mean cortical thickness, or mean surface area were also included as covariates for respective models. The significance of covariates of interest was assessed using the likelihood ratio test, and FDR correction was used to account for the number of variables tested within each imaging measure (i.e., VOL, CT, SA).

Sex differences in the effects of *C4* GREx on brain and behavioral phenotypes were tested by including an interaction term between *C4A* and *C4B* GREx and sex in statistical models. Follow-up analyses were performed in each sex separately.

The total number of participants in each analysis (i.e., those with complete  $PQ-B_{sym}$ ,  $PQ-B_{sev}$ ,  $PP_{bin}$ , phenotype, and/or brain-imaging data) is provided in Additional file 1.

Plots were generated using R packages *ggpubr*, *ggplot*, and *ggesg*. PheWAS results were plotted in R by modifying publicly available code from Dr. Yoonjung Joo ([https://rpubs.com/hellojjoo/pheWAS\\_connectome](https://rpubs.com/hellojjoo/pheWAS_connectome)).

### Replication in the UK Biobank

The UK Biobank resource was accessed through application number 38029. Brain imaging data was acquired for 18,361 unrelated individuals with European ancestry. Haplotypes at the *C4* locus were imputed as described for the ABCD dataset. After further restricting to individuals with the five most common *C4* haplotypes (AL, AL-AL, AL-BL, AL-BS, BS), a total of 16,147 individuals remained in the final analyses. A linear model was then fitted between entorhinal cortex surface area and *C4A* GREx, using *C4B* GREx, sex, age at scan, average cortex surface area, location of head in the scanner (data fields 25756 to 25759), first four genotype PCs, and scan sites as covariates. A null model was also fitted with only covariates but not *C4A* GREx. The statistical significance of the association was evaluated by performing a likelihood ratio test. The analysis was performed in females ( $N = 8357$ ) and males ( $N = 7790$ ) separately, without using sex as a covariate.

### Validation of expression weights for *C4* genes in PsychENCODE

Previously harmonized genotype array and frontal cortex RNA-seq data from PsychENCODE [19, 74] consisted of uniformly processed data from six studies: BipSeq, LIBD\_sControl, CMC\_HBCC, CommonMind, BrainGVEX, and UCLA-ASD. We imputed *C4* alleles in six studies from PsychENCODE separately in the same manner as the ABCD dataset using Beagle4.1 with the latest multi-ancestry *C4* imputation reference panel. This analysis consisted of 340 AFR, 863 EUR, and 50 Latinx samples. *C4* expression was then predicted using previously described expression weights [11], which was associated with observed *C4* gene expression (Additional file 2: Fig. S1). Observed *C4* gene expression was strongly associated with corresponding predicted gene expression, except for *C4B* gene expression in AFR samples ( $P > 0.05$ ); *C4A/C4B* expression weights accurately predicted childhood and adolescent *C4A* expression in EUR PsychENCODE samples between the ages of 5-15 ( $P < 0.05$ ).

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-023-02878-0>.

**Additional file 1: Table S1.** Subject characteristics as assessed at the baseline timepoint in youth with the 5 most common *C4* structural haplotypes. **Table S2.** Count and proportion of *C4* alleles in the ABCD cohort (1,694 AFR; 5,829 EUR; 2,115 Latinx). **Table S3.** Summary of PheWAS results in the ABCD cohort. **Table S4.** Summary of 1-year follow-up PheWAS in the ABCD cohort. **Table S5.** Summary of 2-year follow-up PheWAS in the ABCD cohort. **Table S6.** Summary of *C4* GREx associations with regional brain surface area in the ABCD cohort. **Table S7.** Summary of *C4* GREx associations with whole brain volume (WBV), mean surface area (SA), and mean cortical thickness (CT). **Table S8.** Summary of *C4* GREx associations with regional brain cortical thickness in the ABCD cohort. **Table S9.** Summary of *C4* GREx associations with subcortical brain volume in the ABCD cohort. **Table S10.** Summary of Wilcoxon Rank Sum Tests comparing PQ-B symptoms and PQ-B severity at 1- and 2-year follow-up timepoints in youth in the top/bottom quartile for entorhinal cortex surface area at baseline. **Table S11.** Summary of schizophrenia PRS associations in youth of EUR ancestry. **Table S12.** Summary of schizophrenia PRS associations with regional cortical thickness in youth of EUR ancestry. **Table S13.** Summary of schizophrenia PRS associations with regional surface area in youth of EUR ancestry. **Table S14.** Summary of schizophrenia PRS associations with subcortical volume in youth of EUR ancestry.

**Additional file 2: Fig. S1.** Relationship between predicted *C4A/C4B* expression and observed *C4A/C4B* expression across ancestries. **Fig. S2.** *C4* descriptives by ancestry. **Fig. S3.** Phenome-wide association between predicted *C4A* gene expression and behavioral phenotypes in youth of AFR ancestry. **Fig. S4.** Phenome-wide association between predicted *C4A* gene expression and behavioral phenotypes in youth of Latinx ancestry. **Fig. S5.** Phenome-wide association between predicted *C4A* gene expression and behavioral phenotypes in youth of EUR ancestry. **Fig. S6.** Sex-differences in the association between predicted *C4A* gene expression and behavioral phenotypes in the ABCD cohort. **Fig. S7.** Phenome-wide association between predicted *C4A* gene expression and behavioral phenotypes in female youth. **Fig. S8.** Phenome-wide association between predicted *C4A* gene expression and behavioral



phenotypes in male youth. **Fig. S9.** Phenome-wide association between predicted C4A gene expression and behavioral phenotypes in the ABCD cohort at the 1-year follow-up. **Fig. S10.** Phenome-wide association between predicted C4A gene expression and behavioral phenotypes in female youth at the 1-year follow-up. **Fig. S11.** Phenome-wide association between predicted C4A gene expression and behavioral phenotypes in male youth at the 1-year follow-up. **Fig. S12.** Phenome-wide association between predicted C4A gene expression and behavioral phenotypes in the ABCD cohort at the 2-year follow-up. **Fig. S13.** Phenome-wide association between predicted C4A gene expression and behavioral phenotypes in female youth at the 2-year follow-up. **Fig. S14.** Phenome-wide association between predicted C4A gene expression and behavioral phenotypes in male youth at the 2-year follow-up. **Fig. S15.** Relationship between C4 haplotypes and baseline psychosis-like experiences in the multi-ancestry ABCD sample. **Fig. S16.** Factors influencing entorhinal cortex surface area at baseline.

**Additional file 3.** Review history.

### Acknowledgements

Not applicable.

### Review history

The review history is available as Additional file 3.

### Peer review information

Veronique van den Bergh was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Authors' contributions

This study was designed by LMH and MJG. SYB, CCF, and WKT contributed to the ongoing project design and management of the Adolescent Brain Cognitive Development Study. LMH performed the data analysis, with support from MK, PZ, CCF, WKT, and MJG. LMH, MK, PZ, RAI, GH, RL, DS, SYB, CCF, CEB, WKT, and MJG contributed to the data interpretation. LMH, MK, PZ, and MJG wrote the manuscript, with critical input from all authors. The authors read and approved the final manuscript.

### Authors' Twitter handles

@LM\_\_Hernandez (Leanna M. Hernandez); @mikejg84 (Michael J. Gandal)

### Funding

This work was supported by the National Institute of Mental Health (R01MH121521 to MJG; R01MH123922 to MJG; P50HD103557 to MJG, R01MH107250, and U01MH081902 to CEB; K00MH119663 to LMH; R01MH120025 to CCF; T32MH073526 and F30MH125523 to MK), SFARI Bridge to Independence Award to MJG, National Center for Advancing Translational Science (UL1TR001881 to GDH), National Institute of General Medical Sciences (T32GM008042 to MK), UCLA Friends of the Semel Institute Research Scholar Award to GDH, Karen Seykora NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation to GDH, and the Harvey L. and Maud C. Sorensen Foundation Fellowship to GDH.

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development<sup>SM</sup> (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9–10 and follow them over 10 years into early adulthood. The ABCD Study<sup>®</sup> is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, and U24DA041147. A full list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of participating sites and a complete listing of the study investigators can be found at [https://abcdstudy.org/consortium\\_members/](https://abcdstudy.org/consortium_members/). ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

### Availability of data and materials

The datasets supporting the conclusions of this article are available through the Adolescent Brain Cognitive Development study [21] on the National Data Archive (<https://doi.org/10.15154/1519007>) and with application to the UK Biobank (<https://www.ukbiobank.ac.uk/>) [35]. Genotype array and RNA sequencing data were obtained from PsychENCODE (<https://doi.org/10.7303/syn12080241>) [74]. The codes used to perform statistical analyses are available under the MIT license on GitHub (<https://github.com/leamhernandez/C4A-ABCD>) [75] and have also been deposited in Zenodo (<https://doi.org/10.5281/zenodo.7542574>) [76].

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. <sup>2</sup>Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. <sup>3</sup>Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. <sup>4</sup>University of Cambridge, Department of Psychiatry, Cambridge Biomedical Campus, Cambridge CB2 0SZ, UK. <sup>5</sup>Population Neuroscience and Genetics Lab, University of California, San Diego, San Diego, CA 92093, USA. <sup>6</sup>Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. <sup>7</sup>Lifespan Brain Institute at Penn Med and the Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>8</sup>Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.

Received: 20 April 2022 Accepted: 15 February 2023

Published online: 07 March 2023

**References**

- Klosterkötter J, Hellmich M, Steinmeyer EM, Schultze-Lutter F. Diagnosing schizophrenia in the initial prodromal phase. *Arch Gen Psychiatry*. 2001;58(2):158–64.
- Poulton R, Caspi A, Moffitt TE, Cannon M, Murray R, Harrington H. Children's self-reported psychotic symptoms and adult schizophreniform disorder: a 15-year longitudinal study. *Arch Gen Psychiatry*. 2000;57(11):1053–8.
- Healy C, Brannigan R, Dooley N, Coughlan H, Clarke M, Kelleher I, et al. Childhood and adolescent psychotic experiences and risk of mental disorder: a systematic review and meta-analysis. *Psychol Med*. 2019;49(10):1589–99.
- Feinberg I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J Psychiatr Res*. 1982;17(4):319–34.
- McGlashan T, Hoffman R. Schizophrenia as a disorder of developmentally reduced synaptic connectivity. *Arch Gen Psychiatry*. 2000;57(7):637–48.
- Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460(7256):744–7.
- Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748–52.
- Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–7.
- Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43(10):969–76.
- Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. 2018;50(3):381–9.
- Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;530:177–83.
- Janeway CA, Travers P, Walport M, Shlomchik MJ. The complement system and innate immunity. In: *Immunobiology: the immune system in health and disease*. 5th ed. New York: Garland Science; 2001.
- Stevens B, Allen NJ, Vazquez LE, Howell GR, Christopherson KS, Nouri N, et al. The classical complement cascade mediates CNS synapse elimination. *Cell*. 2007;131(6):1164–78.
- Schafer DP, Lehrman EK, Kautzman AG, Koyama R, Mardinly AR, Yamasaki R, et al. Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron*. 2012;74(4):691–705.
- Comer AL, Jinadasa TI, Sriram BI, Phadke RA, Kretsge LN, Nguyen TPH, et al. Increased expression of schizophrenia-associated gene C4 leads to hypoconnectivity of prefrontal cortex and reduced social interaction. *PLoS Biol*. 2020;18(1):e3000604.
- Yilmaz M, Yalcin E, Presumey J, Aw E, Ma M, Whelan CW, et al. Overexpression of schizophrenia susceptibility factor human complement C4A promotes excessive synaptic loss and behavioral changes in mice. *Nat Neurosci*. 2021;24(2):214–24.
- Dangel AW, Mendoza AR, Baker BJ, Daniel CM, Carroll MC, Wu L-C, et al. The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics*. 1994;40(6):425–36.
- Kamitaki N, Sekar A, Handsaker RE, de Rivera H, Tooley K, Morris DL, et al. Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature*. 2020;582(7813):577–81.
- Kim M, Haney JR, Zhang P, Hernandez LM, Wang L, Perez-Cano L, et al. Brain gene co-expression networks link complement signaling with convergent synaptic pathology in schizophrenia. *Nat Neurosci*. 2021;24(6):799–809.
- O'Connell KS, Sonderby IE, Frei O, van der Meer D, Athanasiu L, Smeland OB, et al. Association between complement component 4A expression, cognitive performance and brain imaging measures in UK Biobank. *Psychol Med*. 2021;52(15):1–11.
- Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, et al. The conception of the ABCD study: from substance use to a broad NIH collaboration. *Dev Cognit Neurosci*. 2018;32:4–7. <https://doi.org/10.15154/1519007>.
- Garavan H, Bartsch H, Conway K, Decastro A, Goldstein RZ, Heeringa S, et al. Recruiting the ABCD sample: design considerations and procedures. *Dev Cognit Neurosci*. 2018;32:16–22.
- Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41(4):1149–60.

24. Karcher NR, Barch DM, Avenevoli S, Savill M, Huber RS, Simon TJ, et al. Assessment of the Prodromal Questionnaire-Brief Child Version for measurement of self-reported psychoticlike experiences in childhood supplemental content. *JAMA Psychiatry*. 2018;75(8):853–61.
25. Kaufman J, Birmaher B, Axelson D, Perepletchikova F, Brent D, Ryan N. Kiddie schedule for affective disorders and schizophrenia present and lifetime version 2013. Working Draft (KSADS-PL); 2013.
26. Cannon TD, Yu C, Addington J, Bearden CE, Cadenhead KS, Cornblatt BA, et al. An individualized risk calculator for research in prodromal psychosis. 2016;173(10):980–8.
27. Lynam DR, Smith GT, Whiteside SP, Cyders MA. The UPPS-P: Assessing five personality pathways to impulsive behavior. West Lafayette, IN: Purdue University. 2006;10.
28. Athanasiu L, Giddaluru S, Fernandes C, Christoforou A, Reinvang I, Lundervold AJ, et al. A genetic association study of CSMD1 and CSMD2 with cognitive function. *Brain Behav Immun*. 2017 Mar;1(61):209–16.
29. Donohoe G, Walters J, Hargreaves A, Rose EJ, Morris DW, Fahey C, et al. Neuropsychological effects of the CSMD1 genome-wide associated schizophrenia risk variant rs10503253. *Genes Brain Behav*. 2013;12(2):203–9.
30. Zhang C, Lv Q, Fan W, Tang W, Yi Z. Influence of CFH gene on symptom severity of schizophrenia. *Neuropsychiatr Dis Treat*. 2017;13:697.
31. Cannon TD, Chung Y, He G, Sun D, Jacobson A, van Erp TGM, et al. Progressive reduction in cortical thickness as psychosis develops: a multisite longitudinal neuroimaging study of youth at elevated clinical risk. *Biol Psychiatry*. 2015;77(2):147–57.
32. Chung Y, Haut KM, He G, van Erp TGM, McEwen S, Addington J, et al. Ventricular enlargement and progressive reduction of cortical gray matter are linked in prodromal youth who develop psychosis. *Schizophr Res*. 2017;189:169–74.
33. Chung Y, Allswede D, Addington J, Bearden CE, Cadenhead K, Cornblatt B, et al. Cortical abnormalities in youth at clinical high-risk for psychosis: findings from the NAPLS2 cohort. *NeuroImage: Clin*. 2019;23:101862.
34. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 2006;31(3):968–80.
35. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–9.
36. Prasad KMR, Patel AR, Muddasani S, Sweeney J, Keshavan MS. The entorhinal cortex in first-episode psychotic disorders: a structural magnetic resonance imaging study. *Am J Psychiatry*. 2004;161(9):1612–9.
37. Roalf DR, Quarmley M, Calkins ME, Satterthwaite TD, Ruparel K, Elliott MA, et al. Temporal lobe volume decrements in psychosis spectrum youths. *Schizophr Bull*. 2017;43(3):601–10.
38. Olde Loohuis LM, Mennigen E, Ori APS, Perkins D, Robinson E, Addington J, et al. Genetic and clinical analyses of psychosis spectrum symptoms in a large multiethnic youth cohort reveal significant link with ADHD. *Transl Psychiatry*. 2021;11(1):1–10.
39. Loughnan R, Palmer CE, Makowski C, Thompson WK, Barch DM, Jernigan TL, et al. Unique prediction of developmental psychopathology from genetic and familial risk. *J Child Psychol Psychiatry*. 2022. <https://doi.org/10.1111/jcpp.13649>.
40. Karcher NR, Paul SE, Johnson EC, Hatoum AS, Baranger DA, Agrawal A, et al. Psychotic-like experiences and polygenic liability in the ABCD Study<sup>®</sup>. *Biol Psych: Cogn Neurosci Neuroimaging*. 2022;7(1):45–55.
41. Howard DM, Adams MJ, Shiralilari M, Clarke T-K, Marioni RE, Davies G, et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun*. 2018;9(1):1470.
42. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668–81.
43. Ruderfer DM, Fanous AH, Ripke S, McQuillan A, Amdur RL, Gejman PV, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry*. 2013;19(9):1017–24.
44. Glanville KP, Coleman JRI, Hanscombe KB, Euesden J, Choi SW, Purves KL, et al. Classical human leukocyte antigen alleles and C4 haplotypes are not significantly associated with depression. *Biol Psychiatry*. 2020;87(5):419–30.
45. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet*. 2021;53(6):817–29.
46. Arnold SE. The medial temporal lobe in schizophrenia. *J Neuropsychiatry Clin Neurosci*. 2006;9(3):460–70.
47. Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. *Schizophr Res*. 2001;49(1–2):1–52.
48. Bakhshi K, Chance SA. The neuropathology of schizophrenia: a selective review of past studies and emerging themes in brain structure and cytoarchitecture. *Neuroscience*. 2015;303:82–102.
49. Honea R, Crow TJ, Passingham D, Mackay CE. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry*. 2005;162(12):2233–45.
50. Mathew I, Gardin TM, Tandon N, Eack S, Francis AN, Seidman LJ, et al. Medial temporal lobe structures and hippocampal subfields in psychotic disorders: findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) study. *JAMA Psychiatry*. 2014;71(7):769–77.
51. van Erp TGM, Hibar DP, Rasmussen JM, Glahn DC, Pearlson GD, Andreassen OA, et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol Psychiatry*. 2016;21(4):547–53.
52. Jakob H, Beckmann H. Prenatal developmental disturbances in the limbic allocortex in schizophrenics. *J Neural Transm*. 1986;65(3–4):303–26.
53. Arnold SE, Hyman BT, Van Hoesen GW, Damasio AR. Some cytoarchitectural abnormalities of the entorhinal cortex in schizophrenia. *Arch Gen Psychiatry*. 1991;48(7):625–32.
54. Arnold SE, Ruschinsky DD, Han LY. Further evidence of abnormal cytoarchitecture of the entorhinal cortex in schizophrenia using spatial point pattern analyses. *Biol Psychiatry*. 1997;42(8):639–47.
55. Akil M, Lewis DA. Cytoarchitecture of the entorhinal cortex in schizophrenia. *Brief Reports*. 1997;154(7):1010–2.
56. Krimer LS, Herman MM, Saunders RC, Boyd JC, Hyde TM, Carter JM, et al. A qualitative and quantitative analysis of the entorhinal cortex in schizophrenia. *Cereb Cortex*. 1997;7(8):732–9.

57. Muñoz M, Insausti R. Cortical efferents of the entorhinal cortex and the adjacent parahippocampal region in the monkey (*Macaca fascicularis*). *Eur J Neurosci*. 2005;22(6):1368–88.
58. Schröder TN, Haak KV, Jimenez NIZ, Beckmann CF, Doeller CF. Functional topography of the human entorhinal cortex. *eLife*. 2015;4:e06738.
59. Ranganath C, Ritchey M. Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci*. 2012;13(10):713–26.
60. Garcia AD, Buffalo EA. Anatomy and function of the primate entorhinal cortex. *Annu Rev Vis Sci*. 2020;6:411–32.
61. Eichenbaum H, Yonelinas AP, Ranganath C. The medial temporal lobe and recognition memory. *Annu Rev Neurosci*. 2007;30:123–52.
62. Moser EI, Moser M-B. Grid cells and neural coding in high-end cortices. *Neuron*. 2013;80(3):765–74.
63. van Haren NEM, Schnack HG, Cahn W, van den Heuvel MP, Lepage C, Collins L, et al. Changes in cortical thickness during the course of illness in schizophrenia. *Arch Gen Psychiatry*. 2011;68(9):871–80.
64. Gilmore JH, Knickmeyer RC, Gao W. Imaging structural and functional brain development in early childhood. *Nat Rev Neurosci*. 2018;19(3):123–37.
65. Tamnes CK, Herting MM, Goddings A-L, Meuwese R, Blakemore S-J, Dahl RE, et al. Development of the cerebral cortex across adolescence: a multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. *J Neurosci*. 2017;37(12):3402–12.
66. O'Brien HE, Hannon E, Hill MJ, Toste CC, Robertson MJ, Morgan JE, et al. Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol*. 2018;19(1):1–13.
67. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. RICOPIIL: Rapid Imputation for COnsortias PipeLine. *Bioinformatics*. 2020;36(3):930–3.
68. Browning BL, Browning SR. Genotype imputation with millions of reference samples. *Am J Hum Genet*. 2016;98(1):116–26.
69. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284–7.
70. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
71. Grasby KL, Jahanshad N, Painter JN, Colodro-Conde L, Bralten J, Hibar DP, et al. The genetic architecture of the human cerebral cortex. *Science*. 2020;367(6484).
72. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature. Communications*. 2019;10(1).
73. Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev Cognit Neurosci*. 2018;32:43–54.
74. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):eaat8127. <https://doi.org/10.7303/syn12080241>.
75. Hernandez LM, Kim M, Zhang P, Bethlehem R, Hoftman G, Loughnan R, Smith D, Bookheimer SY, Fan CC, Bearden CE, Thompson W, Gandal M. Multi-ancestry phenome-wide association of complement component 4 variation with psychiatric and brain phenotypes in youth. *GitHub*. (2023). <https://github.com/leamhernandez/C4A-ABCD>.
76. Hernandez LM, Kim M, Zhang P, Bethlehem R, Hoftman G, Loughnan R, Smith D, Bookheimer SY, Fan CC, Bearden CE, Thompson W, Gandal M. Multi-ancestry phenome-wide association of complement component 4 variation with psychiatric and brain phenotypes in youth. *Zenodo*. (2023). <https://doi.org/10.5281/zenodo.7542574>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

